DOCUMENT RESUME

ED 045 959 AL 002 500

AUTHOR	Hays, Harriet Nutt
TITLE	From Mouth to Hand: Obstacles in Rendering Verbal Events Faithfully into Standard Orthography. The Classroom Interaction Project Series.
INSTITUTION	Missouri Univ., Columbia. Center for Research in Social Behavior.
REPORT NO	TR-CI-40
PUB DATE	24 Jul 70
NOTE	35p.; Paper presented at the Linguistic Society of America Annual Meeting, Columbus, Ohio, July 24, 1969
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Experiments, Graphemes, Interaction, Interaction Process Analysis, *Orthographic Symbols, *Phonetic Transcription, Statistical Studies, Tape Recordings, Video Tape Recordings

ABSTRACT

This paper reports the initial phase of a series of experiments conducted on a large number of videotapes made for the purpose of analyzing public-school classroom interaction. The experiments originally aimed to preduct the most reliable, efficient and economic way of producing transcriptions which are sufficiently representative of the verbal events used for empirical research. Results tabulated thus far indicate that, particularly among nonlinguists, but also among linguists, transcriptions of the same event into standard orthography are apt to differ to a significant extent; that some of these differences may not be entirely predictable; and that it takes at least two iterations of post-editing of the transcript to get a reasonable orthographic representation of the event. It also appears that the more complicated the structures involved, whether they be social, semantic or grammatical, the more verifications or post-editings are needed to produce an accurate transcription. The optimum work increment, processor personality, training or sequencing is not yet determinable, but, especially for difficult passages, it is likely that pairs of judges in the final editings, working together with transcript and tape, will be more efficient than single judges left alone with their idiosyncratic prejudices, anticipations, hearing and experience. (Author/AMM)

ED 045959

CENTER FOR RESEARCH IN SOCIAL BEHAVIOR

University of Missouri-Columbia

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Technical Report #CI 40
The Classroom Interaction Project Series

FROM MOUTH TO HAND:

Obstacles in Rendering Verbal Events Faithfully into Standard Orthography

AL 002 500

by

Harriett Nutt Hays

From Mouth to Hand: Obstacles in rendering verbal events faithfully into standard orthography

(Paper presented at the annual meeting of the Linguistics Society of America, July 24, 1970, at Columbus, Ohio, by Harriett Nutt Hays, Research Associate, Center for Research in Social Behavior, University of Missouri, Columbia, Missouri.)

To a phonetician or a lexicologist who has had direct experience in the field with variations among observers who transcribe or analyze oral data, either live or from electronic recordings, it is a matter of fact that all observers do not hear the same utterance in the same way. Kurath and McDavid, in Linguistic Atlas discussions, have reported variations among trained field workers. The Swedish dialectician Ringaard found that perceptions of other dialects by trained phonologists were influenced by their own manner of speech. Lieberman has noted that a great deal of the linguist's perception of prosodic features is based on intuition or knowledge of grammatical structures. I. H. Paul, a psychologist, noted some time ago that listeners' recall of a verbal event was subject to considerable variation among observers. Recently, Gumperz has bemoaned the fact that it is very tedious to obtain a transcript of an oral event for sociolinguistic studies.

What emphasis there has been on observer discrepancies has generally concentrated on subwordal, or phonological phenomena, or on more abstract functional or semantic entities. Very little has been done to examine the actual extent and cause of observer discrepancies relative to the translation of an oral situation into standard orthography. The assumption in many fields seems to be that a written transcription will not vary extensively from the real event if it concentrates on representing just the uttered words and sentences of that event: it is the prosodic information which is apt to be distorted. As we all have observed, transcriptions of wordal events are sometimes used as primary data for that event, for research, for legal action, for political record.

The clue to the general ignorance of the difficulties involved in capturing what is actually said may be reflected in the problems of illustrating and assessing them. It is, indeed, extremely tedious to represent oral information in a permanent form which is easily accessible

for perusal, and it is even more tedious to analyze varying representations of an event.

Signals made simultaneously in several dimensions of one medium have to be telescoped into perhaps only one or two dimensions of another. What is represented, for instance, in speech by the quantitative indicators (prosodic features) of amplitude, frequency, rate, and duration, which occur simultaneously with the qualitative indicators (consonants and vowels) are represented in writing as two-dimensional graphic signs (punctuations marks) which usually appear sequentially to the qualitative symbols (letters). The only simultaneous indicators for standard formal orthographic representations are capital letters, italics, underlines, boldface print, and the like. There is not a one to one ratio of the two symbolic systems, so their confusion is inherent in any transliteration. This is then compounded by the requirements of written discourse that all utterances be segmented by terminals which enclose strings of supposedly specific structures including what are referred to in standard school grammars as 'subjects,' 'predicates' and 'complete thoughts.' Spoken discourse, particularly with informal style, is characterized by what would be considered 'fragmentation' in written grammatical tradition. The problem of representing these 'incomplete thoughts' is difficult for translators who have been given no guidelines, particularly when they may have differing views of 'completeness' or of 'grammaticality.'

There are other problems which accompany the conversion procedure. Both the media and the situations for production of speech and production of its transliteration are different, presenting numerous possibilities for distortion. Speech, which can be considered the primary data, is a relatively unpredictable string of events incorporating a number of mutually interactive semiotic systems among which oral gesture and other expressive behavior patterns are included. Transliteration of the speech event is much more restricted in its potential boundaries than is the production of that event. Yet in both situations the participants have to add their individual interpretations to the events. Both are translation systems of a sort: speech is the translation into sound, apparently of thought product or behavioral convention; transliteration is a translation of a part of the speech event into a graphic medium. Just as speech suffers from limitations in its signals which do not

represent all the elaborations of the mind, so transliteration suffers from a lack of signals which simultaneously reinforce, extend or contradict nuances of the verbal message. The system which the transliterator uses seems rather to be designed to create graphic events such as essays, novels, poems or letters than to convert an alien system to its form.

The transliterator has the freedom of neither the author nor the speaker. Like the hearer and the reader, he must bring to the communication system his own experience in order to interpret the multi-referential signals which are employed. But unlike the hearer and the reader, he may not include them overtly in his transcription. His is a translation problem, in which a part of the other code is systematically left out. The results of such a conversion, without supplementation of electronic recordings of the oral event may be quite unusable as primary data, even though they may be more valid than the recall of the event by any one individual or individuals who do not commit the recall to a transcription.

This paper reports the initial phase of a series of experiments conducted on a large number of videotapes made for the purpose of analyzing public school classroom interaction. The original aim of the experiments was to predict the most reliable, efficient and economic way of producing transcriptions which are sufficiently representative of the verbal events to be used for empirical research.

The results of work tabulated thus far indicate that, particularly among non-linguists, but also among linguists, transcriptions of the same event into standard orthography are apt to differ to a significant extent, that some of these differences may not be entirely predictable, and that it takes at least two iterations of post-editing of the transcript to get a reasonable orthographic representation of the event. It appears also that the more complicated are the structures involved, whether they be social, semantic or grammatical, the more verifications or post-editings are needed to produce an accurate transcription. The optimum work increment, processor personality, training or sequencing is not yet determinable, but, especially for difficult passages, it is likely that pairs of judges in the final editings, working together with transcript and tape, will be more efficient than single judges left alone with their idiosyncratic prejudices, anticipations, hearing and experience.

The following discussion illustrates the kinds of omissions or distortions which the transliterator is apt to make when he transcribes an oral event using the standard literary graphic code. Note that the discrepancies tabulated refer only to the differences within the standard graphic transliteration system. They do not take account of the real situation, the actual amount of information of the communication system which is preserved or lost because of deficiencies in the system or elsewhere.

In order to assess the results, of course, it was necessary to use a preserved oral event. It was not possible at the time to make recordings specifically for this purpose and to have multiple observers on the spot judge the relative fidelity of the recorded material to the live situation. Such information was not extant for the recordings which were available, so no judgments about their actual fidelity to the live situation will be pertinent for this study.

It is pertinent, however, to know the conditions under which the recordings were made. Both segments of tape discussed in this study were recorded on the same 2400 foot roll of 2 inch 3M videotape, in the same urban elementary school, one each from two different sixth grade classes. Both of the teachers of these classes were young (20-30) females, who appear to speak the standard (prestige) teacher dialect of that Missouri city. The teacher of Segment I was black, of II was white. Class I was in English composition. Class II was in Social Studies, apparently a geography lesson in which a certain amount of reading aloud from the textbook took place. The pupils were male and female, black and white, children of approximately 12 years of age who spoke the local urban dialect but did not appear to have adopted so-called standard American.

Six microphones were used. The teacher wore a microphone suspended about the neck. The audio channel from this microphone was recorded on one track of a two track recording system. Four other microphones were hung from the ceiling and were recorded on the other track. All these were supplemented by another, directional, microphone which was aimed at the immediate emitter source. Two cameras were installed at opposite points in the classroom. The teacher was kept in constant focus, and the teacher picture inserted into an unused portion of the general classroom picture.

A detailed description of the recording equipment is found in Biddle and Adams, 1967. The recordings used for this study were made in 1968, for the Center for Research in Social Behavior at the University of Missouri, Columbia, Missouri. There was no special attempt made by teacher or pupils to enunciate or otherwise distort their behavior in order to improve the video and audio clarity of the recording. There were occasions of single as well as multiple responses by members of the class. The teacher was usually audible, but sometimes members of the class stationed far from the general microphones were difficult if not impossible to distinguish. Sometimes part of the class was not visible on the videoscreen. The lighting (unsupplemented on a rainy day), the distance, and the focus of the recording were such that the facial expressions of the class members were often not perceptible. Although a seating chart and rosters had been obtained at the time of recording, apparently there was no check on the actual position or presence of individual pupils, for neither these nor other tapes in the series reflect very well the arrangements indicated. (On some tapes there is no relationship at all between the rosters and the arrangement or content of the class). As we will see, this proves unfortunate, and, for those researchers for whom it is not a matter of course to diagram recorded events for location and activity of participants it would be well to take note. Literally hours of weeks have been spent by us attempting to straighten out boys from girls, black faces from white ones, high voices from low, etc., with very unsatisfactory results. It is our experience that, if the information is not gathered at time of collection, it might well be permanently lost. This loss of essential informant information then limits the use to which otherwise acceptable materials might be put.

The tape segments were arbitrarily chosen, with no pre-examination of the tape itself, for the contrasts originally were to serve as a quick illustration to associates that caution needed to be exercised in interpreting the oral material. The transcribers at the time were engaged in transcribing the tapes for the recorded series, and the next tape scheduled for transcription was selected for analysis.

Both transcribers and editors used the same playback equipment: a standard CONRAC 240" monitor set, an AMPEX VR01500 portable tape recorder with Play, Fast-Forward and Reverse controls handling two channels and accommodating a Tandberg 22 footpedal with two button controls for Forward and Reverse. A Sharp headset connected to a switch-box for signals from either channel singly, or both channels stereophonically, to the earphones, was used by some transcribers and editors. Others preferred to listen without earphones. No attempt was made to standardize or control this, but a casual survey indicates that it is probably more efficient to use the earphones, which seem to cut off transcription environment interference noise. Some persons, however, complained of headache from the headsets, and it was assumed that for them it was more efficient to work without both headache and headsets.

A survey of business concerns regarding optimal work increment for persons operating dictaphones seemed to indicate that a twenty minute period might be optimal for transcribing. This was shocking to the secretaries involved who had been in the habit of spending a much longer increment. A compromise of about 40 minutes was finally settled upon for transcription sessions. Time allowable for paper insert, forward and reversing the tapes and examination of the video image were assumed to constitute proper rest periods within the 40 minutes. (Examination of manuscripts seems to indicate that transcriber efficiency decreases rapidly at the end of 60 minutes.) About half of the editors claimed that they didn't start getting efficient until they had been working for about an hour, so there is certainly divided opinion on work increment. Some of the editors also claimed they could do better transcriptions than the transcribers by working at long intervals by hand. (This has been a common reaction. Most persons who have seen the lists of discrepancies have volunteered the information that they themselves would not make errors.) Since none of them turned out to be infallible, it is still unclear whether relative time has any bearing on the quality of output. There is no question about the fact that it is much more expensive to employ a slow working than a fast working editor.

Editor is really a misnomer. The editors were not to function as standard literary editors do. Rather they were to reproduce an utterance without improving upon it. They were to compare all previous graphic versions against the tape of the situation and to insert, delete or modify

those translations which seemed inappropriate. Similarly, the transcriber was to reproduce rather than beautify the original utterance.

The regular transcribers for the interaction study then, four secretaries employed by the research center, one a college graduate, the others with high school degrees and secretarial training, transcribed Class I using the conventions they had already established: All senders of utterances were designated in parentheses, left-justified on a new line. All public utterances were transcribed into standard orthography. Indistinguishable utterances were represented by a line, whose length might or might not have an impressionistic relationship to length of utterance.

Class II was transcribed by the same set of secretaries but after a two hour session to establish additional conventions. Annotations or disambiguations to clarify the context of ambiguous utterances were to be inserted in slashes, on the assumption that, if they were properly marked they could easily be left out. Punctuation was to be reduced to a minimum, where possible. For terminals, only period, indicating a statement neutral in tone or feeling, question mark to indicate a definite question contour, and three dots to mark a suspended or unfinished oral sentence were to be used. Quotes were to enclose matter being read out loud by an emitter. (The later convention ...# used by some editors, had not yet been established. The practice of indicating pauses was also not standardized until the post-editing was well under way.)

The determination of the length of the tape segments contrasted was calculated from the first phrase for which there was consensus among versions to the last event of the shortest transcript, a total, for Class I, of seven minutes of real time. The initial segment, on which this common point was established is represented on page one of the handout. The shortest transcript was that of c, which had one indistinguishable utterance of the first emitter before the common beginning phrase and none at the end of the common transcription. Next was d with two distinguished utterances by the first emitter, before the common starting point, and 4 lines (2 emitters) beyond the end point. B began at the common point, but continued for 9 lines (6 emitters) past the common end. A had one distinguishable utterance before the first, and 11 lines (6 emitters) after the end point of the common transcription.

The corresponding transcription segments were compared for gross characteristics which are often used to describe manuscripts: number of lines, sentences, words, emitter types and word totals. Page and line are obviously inadequate descriptive categories for manuscripts unless they are all made on the same size paper with the same size type. The typed abcd transcripts were, but some of the other transcripts in this study differed in line spacing, size of paper, and size of graphic symbols. (The handwritten copies had either bigger or smaller symbols than those which were typed.)

Number of sentences for the manuscripts was similar, but on close examination the content of the terminals were found to contrast sharply. In the tally for potential sentences, 91 potentials, or 4 more than the greatest number of sentences indicated by any one transcriber were postulated, but might vary for each analyst contrasting the scripts.

The fifth original transcript was made by secretary b, an arbitrary choice, for both Classes. All transcripts but II c and d were then edited by a team of individuals listed on page 8 of the handout.

The first four unedited or raw transcripts for Class I were extensively contrasted with each other for differences in major block categories of Emitter, Annotation, Punctuation and Utterance. These were subdivided further according to a hierarchial code devised to prepare the data for ultimate input into a computer, where the long lists of idiosyncratic and other deviations might be tallied with greater ultimate ease, or at least accuracy.

From the raw transcripts for Class I a handwritten transcription (E) was calculated based upon majority agreement of processors and contextual fit. This calculation was done by a naive editor, that is to say a non-linguist, relatively unsophisticated female white sophomore. It was edited against the tape by the same person, then re-edited by myself and a sophisticated, acute, male white sophomore. The second transcription by b, (F) was edited four times against the tape. An exhaustive chart was drawn to align the manuscripts, and make some hand tabulations, by a linguistically naive female white senior, with apparently good judgment.

For some of the contrasts, more than one chart was drawn and compared. G, m, and n all worked on contrasting some of the versions.

From the final forms a linear alignment was made by g in preparation for entering all of the contrasted information into a machine for tabulation. The extensive coding for this has undergone a number of recent revisions, in consequence of which the machine tabulations for individual variables, processor agreement ratios, etc. has not been completed. It should be in final form shortly and the results prepared for publication, along with the substance of this paper.

The second set of manuscripts were not hand contrasted as extensively as the first, since the element of fatigue and tedium seemed to be so strong for everyone enlisted to tabulate. Last minute hurried hand tabulating sessions in which I enlisted all manner of friends and relatives support this decision. Although there are hand tabulations for most of the blank figures in the statistic sheet on the handout, it was determined not to enclose them because of their probable unreliability.

Because of the revision of the code, the overall key was not included, but this may not have been a wise decision. Suffice for the moment to say that the primary aim in the breakdown and match of each version was to contrast wordal units. Words are defined as strings of graphic symbols unbroken by spaces. Partial words of standard orthography, misspellings, nonsense words are all classified as words. Except for one instance, words are not subdivided into morphological or phonological categories for purposes of analysis, but where discrepancies seem to hinge on phonological criteria, or morphological, or any other, there is stipulation in the code to mark it. These instances then can be extracted and further analyzed by hand or re-entered into the data base.

It will probably be best to discuss the process of aligning, matching, classifying and analyzing the discrepancies from the sample portions in the handout. I should mention that what appears initially to be trivial surface contrastive analysis, shortly approaches the complexity of, and ultimately becomes a problem in discourse analysis. Once an attempt is made to explain the discrepancies, or even to classify them into fine categories, the usual theoretical problems of linguistic string analysis develop. Since part of my contention in oral to graphic data preparation is that the data is apt to be distorted at any point in its handling, there is an attempt to mark processor inference source.

This part of the code will probably undergo still more revision. This may be because it is not always clear what the processor decisions were based on: influence of the manuscript and previous processors, influence of the social situation, influence of invisible cultural knowledge, phonological environment, anticipation of event, dialect interference or overlap, white noise, poor hearing, or whatever.

Please look at page 1 in the handout which gives the unedited transcriber versions for the initial portion of the Class I tape segment. It is interesting to note that this contrast of a,b,c and d is a relatively simple four way job. After each version was edited one or more times and new versions added, it became much more difficult to compute, separate, align and explain them.

Immediately obvious are the gross contrasts of line and word count difference. A has 7 lines, b and c 8, d has 9. Total word count is similar: for a, 67; for b, 70 plus Z; for c, 66 plus Z; and for d, 70. Z in this case indicates indistinguishable utterance. It has proven to be a convenient category, but can contain anywhere from 0 to 10 or 11 potential units (which, so far are the longest strings of universl agreement).

In all, 91 contrasting analysis sections were set up for this sample of text. Note that this is higher than the total word count, but that Emitters, Annotations and Punctuation are excluded from the total count. The number of contrasting analysis sections for the sum of manuscripts covering this tape segment is much greater, since both potential word items and annotation items grew significantly, especially when disambiguating and filling in Z utterances. Punctuation sections increased slightly.

Wherever there is a change of block category among items the analysis procedure calls for a new section to be established, for it has proved convenient not to mix the categories. Occasionally this rule has caused initial inconvenience, but thus far has been extremely useful.

The greatest inconvenience was relative to the initial capital letter of standard orthography which marks sentence initiation. This punctuation category was the last subcategory to be included in the code, and was set up finally as the most convenient way to represent the functional signaling of sentence onset, which, though usually

redundant with a preceding space or terminal, is not always so. There were a number of cases where the only contrast between versions was in this initial capital (The; the). In some cases it was the only similarity in the utterance initial word, as in the case of the discrepancies for (Oh, All, Allright), represented in the text sample and on the second page of the handout. The category can be used in the same way without ambiguity, to mark proper names (grandfather, Grandfather, p. 2).

What I've done is to set up a separate punctuation unit, initial capitalness, which is coded as having no official unit position, but as having a form and functional referent of each one. The processor identification and agreement can then be listed and processors who indicate utterance initial included, even though the form selected for initial position is quite different from that selected by one or more other processors. Originally, it should be noted, this difference was placed under a sub-category of spelling. Since spelling now has been reabsorbed into a word discrepancy subclassification, the new distinction is fortuitous. The mechanical problem of text output is solved by setting up a Partial Word category. Thus the capitalness is marked under one analysis section. The next section represents the actual capital letter, and any small letter, Z or 0, which contrasts with it. The segment classification indicates that it is a partial word and therefore to be joined with the immediately succeeding form, which is represented in the next, separate section. (A rule has been set up that the partial word segments must be the only units in a segment section.)

To review, the hierarchy runs: Section (an analysis convenience for tallying), Block, Class, Segment (phrase, word, partial word), Unit. In some segments there may be as many as eleven (longest universal utterance) items, but the usual number is one. The most convenient ultimate resolution for the machine will probably be sections of one item each. For human beings, however, it is more convenient to put all members of a universally agreed upon string into the same section unless there is some reason for signaling one or more of them out. (I have had particular cause to note the difference in convenience of arrangement of materials prepared for man and for computer. The original text matching was done in charted, columnar form, with major divisions of block and section. Universal agreements were itemized separately from discrepancies.

When this was rearranged back into linear form, each section containing the matched versions, it was far more difficult to hand tally the results.

Subclassifications for Emitter are Multi-person vs. Single person utterance. This represents one of the most difficult problems: representing utterance overlap in a linear string. The problem is not yet solved to satisfaction for text or script output, but can always be handled by annotation. Annotations are divided into Behavioral, Locational, Timing, and Prosodic. There can be overlap or multiple classification of any of these. This is true of all subcategories unless a specific prohibition is made. Primary categories for Punctuation are Initial, Medial and Final. This proves more satisfactory than the original classification of internal and external. Most medial punctuation discrepancies are trivial. Perhaps the most salient are those which represent a short pause which may be important for referential disambiguation. Utterance is divided into standard heuristic formal grammatical categories based upon the first member of the section, irregardless of the grammatical structure represented. It is assumed that sections needing finer classification will be extracted and re-run. The sub-categorization system is currently loose enough so that additions or refinements to subcategories can be made before input. The data is so complicated, however, that revising and updating classifications would be difficult after input.

Universal sections are those on which all processors agree. Discrepant sections are divided three ways: into O for Omission, which includes any contrast having a zero number, that is, an indication of non recognition of the utterance in a manuscript. Z indicates an Indistinguishable utterance, what is represented in the text as a series of dashes or an empty underline. Interpretation is the class term for versions, each of which have a concrete suggestion. For each section, Processor Agreement for Form, Lexical Referent and Functional Referent are coded for

Units, as well as Rank of Probable Significance, and Processor Inference Source. The last two are difficult to assign, and tend to rest more on conjecture than the others. Rank of Significance, obviously will vary depending upon the purpose to which the transcript will be put. For some purposes change in word count will be crucial; in others, referential alteration will be crucial. Yet others will count as trivial or neutral anything but the most glaring case of contrast.

Emitter omission, indistinguishability or mistaken identity is significant for most research. Omission in general may or may not be significant. For this neutral code it is usually ranked as 4, next highest in significance, irregardless of the functional, semantic, formal or other load which it carries. For specific purposes the code will have to be tailored to demands. The code is also loose enough to allow for multiple sub-classification within categories. For text processing this seems to be absolutely necessary since there is so rarely a one to one correspondence between any particular form and referent.

By now you have probably had a chance to read through the text samples and eye match them for yourself. Let me read quickly through the analysis section classification, and point out some exemplary difficulties. Starting with the top transcription, a, note that the first three sections will be devoted to resolving the initial capital 1) (capitalness) 2) (H-) Word Partial, 3) (-ave) Word Partial, all of these, Universal utterances. The following section 4) (all of you) is also UH (for Phrase) with deviation in the succeeding word, 5) (all) vs. zero, 6) (completed the reading) HU, 7) (on/of), 8) (page 109), UH.

For this text (109) is Universal, but for succeeding edited versions it has been convenient to set up two to four units, either as separate sections or as members of the same phrase section. Potentially alternate forms are: one oh nine, one zero nine, one hundred nine, a hundred nine, one hundred and nine, a hundred and nine, hundred and nine, hundred nine: a total of at least nine possible combinations of seven different word forms, into segments of one, two, three or four units. This complicates unit or word count for the separate sections, and could introduce referent ambiguity relative to the rest of the text.

Section 9) (?/.): Terminal Punctuation. 10) (Class): Emitter for b,c,d, but zero for a. 11,12,13) take care of the capital letter: zero contrast as well as the word:zero contrast for (Yes): b,d, and zero for a,c. 14) (Murmurs):c and zero for the others, is an unmarked Annotation Omission. For convenience it is assumed that the period belongs within the annotation, since none of the other manuscripts used punctuation. This may be unrealistic, since a terminal was probably intended by the transcribers who indicated an utterance, but, to save a section, it was not included. Needless to say, in later, edited versions, a punctuation section was introduced at this point.

Section 15) indicates a new Emitter (T) in three of the manuscripts, but zero in a. 16,17,18) take care of the three way contrast of (Allright, Oh/All right). 19) has (right) for three versions, zero for c. This is a Phrase/Word analysis problem analagous to that for numerals: On the one hand the referent is apparently the same for the idiom, but individual words making it up can be analyzed or counted separately. In such cases, the code has been marking the idiomatic components both H and W. Originally the sections were then subdivided into two or more parts. It is beginning to appear, however, that it will be more reflective of the facts if each component is assigned a separate section.

Section 20) is presence or absence of comma. 21) (let's discuss how a): Universal H. 22) (friendly/formal) a glaring Word Interpretation referent discrepancy, c: abd. 23) (letter) UW 24) comma vs. zero, d:abc. 24) (letter) vs. zero: d:abc. 25) (can be like having a visit with someone) an 8 word Universal Phrase. 26) period, a Universal Punctuation Terminal.

Sections 29,30,31) take care of initial capital and (W-) (-hat) Universal Word Partials, 32) (are some of the) 4-word Universal Phrase. 33) presence: d vs. absence: abc, of comma 34) (uh) vs. zero d: abc 35) comma: d, vs. rest: Internal Punctuation Omission. 36) (things) bcd: vs. zero; a Word Omission. 37) (uh/a) cd:ab Word Interpretation. 40) comma vs. zero: d:abc Medial Punctuation Omission. 41) (talk about in a) 4-word UH. 42) (friendly:formal) abd:c Word Interpretation. 43) (letter) UW. 44) (?) UP Final.

Section 44,45,46 cover the universal agreement on sentence initiation and the word involved. 48) continues the universal agreement phrase for 4 words. 49) repeats the friendly/formal discrepancy. 50) gives the Universal Word (letter) and 51) ends this emittance with a Universal Punctuation Terminal: (?)

Section 52) indicates a continuation of Universal agreement, but a new block division: Emitter, Indistinguishable. 53,54) represent the initiation of an utterance, by abd, Z by c. but 55) shows that the word of initiation is different for d than for a and b: c continues with Z, Undistinguishable utterance. Note that 53) gives the Punctuation Initial, agreed upon by abd. 54) has (Y) for ab, (T) for d, Z for c, and is classed as Partial Word Interpretation discrepancy. 55) is Partial Word Omission discrepancy: (ou) for ab: Z for c, and zero for d. 56) covers (mean) vs.

zero vs. z: ab:d:c Word Omission discrepancy. 57) has (T/t) vs Z d:ab:c which completes the sentence initiation begun in 53). 58) gives the rest of the Partial Word Interpretation discrepancy (he). 59) represents the presence of a comma in d, but not in the other manuscripts. 60) (uh) vs. zero vs. A. d:ab:c Word Omission discrepancy: note that when both Omission and Indistinguishability occur, Omission takes priority in classification. 61) gives d's second comma, unrecognized by ab: who have zero, or c, who has Z. 62) is the Universal Word agreement (parts) and 63) the Punctuation Omission discrepancy: question mark vs. period vs. Z. b: ad:c.

Section 64) indicates a new Universal Emitter (T). 65,66,67) the Universal Sentence Initiation and Word agreement. 68) a Medial Punctuation Omission discrepancy: comma for c and d as opposed to zero for a and b. 69) (not the parts) is a Universal Word 70) another comma Omission discrepancy acd:b. 73) UH (the atmosphere) 74) UPF (.)

75) begins one of the most interesting four-way contrasts: it represents the Universal Sentence Initiating capital. 76) gives the Universal initial (D) 77), however, has four Word Partials to complete the initial: (avid) (oyle) (ale) (ana), for a,b,c,d, respectively. If you turn to your handout, page seven, you will notice the discrepancies on apparently similar phonological strings which were tabulated. The apparent choice for this segment was David in the final edited version.

78) gives the Punctuation Terminal Interpretation discrepancy, (?) vs. period. bd:ac. 79) continues the four-way contrast in interpretation by the transcribers: three names vs. one initial. For this contrast the analysis sections did not subdivide emitter or annotations. The final contrast does, so there would be more than one analysis section to account for the information in (79).

80) like 53) gives the Universal Sentence Initiation capital, and 81) like 54) indicates that the capital letters involved are different, having a (Y) for b, and zero for the three others. 82) has the Word Partial for b, the zero for acd, making it an Omission section. 83) is a straight Word Omission with (mean) for b and zero for the others. 84) covers the Partial Word Omission discrepancy of (T/t) a,c,d just beginning their

utterances, and b continuing hers. 85) reveals that all the transcribers agreed on the Partial Word Interpretation (he) 86) continues Universal agreement for 4 words. 87) shows that d and b heard something, that a and c did not: (you/who) vs. Zero for Word Omission and Interpretation discrepancy. Notice again, that where an Omission occurs in a section, the section is marked for it, but amy also be marked for Interpretation if more than one version of an utterance exists. 88) is a Word Omission and Interpretation discrepancy, giving alternate forms of the same verb referent (writes/write/writin) opposed to Zero. a/d/c:b 89) gives a possible pronominal object (it) for c, and zero for the rest. 90) has an interjection (like) for b as apposed to zero for the others. Note that, phonologically, (like) and (it) might have been classed in the same section. They were put into separate sections because of their referential difference and the possibility that both transcribers were correct in distinguishing them. As it happens the final editor chose (like), leaving out both the (you/who) and (writes/write/writin) sections. 91) closes this stretch with a Universal Punctuation Terminal, period.

A summary of statistics gives a total of Emitter blocks for this stretch as 5, with a having four of these realized, and each of the others having five. Within the sections, however, although the segment unit count agrees on one unit per section, there is a choice of eight forms: zero (zero/class) (zero/T), ?, Z, T (one of which was universal), Class, and the four versions of the student name. Processor agreement varied from U, to a four-way discrepancy. Where the agreement ratio was 1:3, b was the dissenting version twice (zero vs. T/Class). In the choice of names, a may have made the correct interpretation.

Page four gives two samples of edited transcripts which are deceptively neat in appearance. Unfortunately both the labels and the first part of the top sample were omitted from the stencil. The first section comes from Fbefgk, and has three colors of ink and different handwritings to code the editors. The bottom sample is Bbmm, with two colors, black and blue, which are sometimes difficult to distinguish. In the part of the stencil

omitted, even the colors fail to distinguish the recommended sequencing clearly, for in a horizontal space of two inches and a vertical space of two lines, an annotation is revised and expanded to require sections for an Emitter, Utterance, Annotation and Punctuation. Arrows drawn in the individual colors serve for additional decoration, but not always clarification.

You will notice that in places of text ambiguity the editorial annotations increase significantly, and, though we have no universal annotations in form, there are a number which agree in substance. The problem here of Emitter two (Mr. Wesley) struggling with the pronunciation of Moslems and Philippines, and Daryl (whoever he be) trying to conquer the word for Portugese, are interesting graphic-phonological discussions. The handout on page seven gives one representation of first syllable spelling, where there seems to be a discrepancy in representing a glottal cut-off. There is another discrepancy on Portuguese which is not illustrated here, but has a seven to nine way choice, depending on how it is sectioned, Interpretations contrasting with both Z and zero.

The easiest way to resolve difficult matching problems has been physical alignment of sentence versions bounded by universal agreements and matching from both ends of the discrepant stretch. This is illustrated on pages five and six. Unfortunately the reference numbers were omitted from the stencilled columns. If you number them for yourself, reading sequentially across the page, you will find twenty-five slots in the first example (part of which inadvertently got placed on page 6) and twenty-two position potentials in the second example. In the second there seem to be an error: columns three and four should be combined.

It is very difficult to match utterances which contain phrase repetitions. It appears that the processors infer phrase repetition, and go ahead to repeat a phrase. The problem is that each processor may settle upon a different phrase of the utterance to repeat. Thus you might postulate in the analysis four repetitions, where only one actually existed. The problem bears examination, and might contain a clue to translation process. I would like to run some studies on observer reliability on a predetermined, canonic text, that is, one which has already been finely edited so that the actual set of repetitions has been clearly analyzed.

One reason for the empty looking statistics sheet, the final page of the handout, is that one or two frequency tabulations mean nothing unless they are carefully matched to processor agreement, referent discrepancy and final choice. As I mentioned above, hand tabulations were made of most of the slots in the matrices, but, because of changing code, resegmentation, and some shift in the analyst outlook, they were withheld from the stencil. When they are all counted by a machine, even though error will still be inherent, they will constitute a more legitimately manipulable set of statistics. For the moment the numbers entered here will give an outline of the parameters.

A note on time spent on the manuscripts: The figures in line 34 are totally unrepresentative, and list only the initial, formal beginning and endings of tasks. They in no way represent the actual number of hours spent on reviewing both manuscripts and tape. It would probably be realistic to assume that, except for the initial stages of transcription and editing, all the rest of the figures should be multiplied by several. Note that there are several ways of calculating work increment for the manuscripts E and G which represent the total of experience before them, plus new judgments, either just the collator and editors of those versions could be considered, or the entire set of previous transcribers and editors listed as persons working on the stretch.

The processor contrast project is still in its infancy. Much still has to be done to establish whether or not there are patterns to discrepancy types. Thus far the only really predictable discrepancies are for medial punctuation (it will be disagreed upon about half the time) and utterances clouded by overlap or outside interference (where each processor will tend to extract different portions of the utterance (see p. 5 and 6). Proper names are bound to be disagreed upon, especially if there is no way to predict their likely form. Vocal gestures such as um, ah are apt to be left out or spelled in varying ways. The spelling could be standardized, the omission is more difficult to combat. There seems to be no definite trend in the samples analyzed to indicate that any other grammatical category is particularly open to disagreement.

It seems unlikely that an algorithm will be formed immediately to calculate most likely choice among discrepancies. I have tried combining processor agreement ratio, referent rank, etc., but come up so

far with nothing satisfactory. Context of situation has to be accounted for, but it is doubtful that this can be automatized yet.

There certainly is evidence that a great many discrepancies among observers occurs. Whether or not these discrepancies are adjudged significant depends upon the use to which the observations are to be put. Where an accurate rendition of events is desired, for instance in a court transcript or a social science investigation emphasizing exactly who said what, most discrepancies will be significant in some way. The ultimate aim of these experiments is to establish editing conventions and agreement procedures so that it will be possible to complete the loop of the translation, from emitter to transcriber to editors to emitter, and get to the heart of the informant utterance.

References

Biddle, B. J. and Adams, R.S. "An Analysis of Classroom Activities." A Final Report, Center for Research in Social Behavior, University of Missouri, Columbia, Missouri, 1967.

Gumperz, A. "Sociolinguistics and Communication in Small Groups" (prepublication draft), University of California, Berkeley, Calif.

Kurath, H. "Handbook of the Linguistic Geography of New England." Providence, Brown University, 1939.

Lieberman, P. "On the Acoustic Basis of the Perception of Intonation by Linguists." Word, 1965, Vol. 21, pp. 40-54.

Paul, I. H. "Studies in Remembering: The Reproduction of Connected and Extended Verbal Material," in Psychological Issues, G.S. Klein, Ed., International Universities Press, Inc., New York, 1959.

Ringaard, K. "The Phonemes of a Dialectal Area, Perceived by Phoneticians and by the Speakers Themselves." Proceedings of the 5th International Congress of Phonetic Sciences (Munster, 1964) 495-501, Basel & N.Y.: S. Karger.

OTHER PAPERS AND REPORTS
by Harriett Hays
Relative to Investigations of Observer Discrepancies
(Produced at the Center for Research in Social Behavior, University of Missouri, Columbia, Missouri)

1. The Need for Post-editing and Standardizing Transcriptions of Classroom Tapes. October, 1968.
2. Transcriber Inconsistency Patterns: Study I. 5 November 1968
3. Problems in Transcriber Training. December, 1968
4. Videotape Operating Procedures: A Working Manual for Transcribers and Editors. February, 1969.
5. Transcribing Procedures: A Working Manual. March, 1969.
6. Cataloguing Procedures and Forms: A Working Manual. February, 1969.
7. Information Retrieval Sources and Suggestions for the Classroom Interaction Project, March, 1969.
8. Problems of Converting Speech Events Reliably into a Graphic Form Compatible to Empirical Research in the Social Sciences. February, 1969. (Delivered to University of Missouri Linguistics Colloquium, December, 1969.)
9. Compilation of the Transcribing Procedures used for Classroom Interaction Videotapes. November 1968-July 1969.
10. Compilation of Post-editing Procedures used for Classroom Interaction Videotape Transcriptions. November 1968-July 1969.
11. From Mouth to Hand: Obstacles in Rendering Verbal Events Faithfully into Standard Orthography. July, 1970. Paper delivered at the Annual Meeting of the Linguistics Society of America, July 24, 1970. Center for Research in Social Behavior Technical Report # 40. July, 1970.

IN PREPARATION

1. Transitions to Terminals, Utterance to Sentence.?....: Discrepancies among judges assigning orthographic sentence boundary markers to transcriptions of natural speech events.
2. David, Doyle, Darla, Daryl, Dana, Dale, Gill, Keel, Gail, Sheila, or Bill? Discrepancies among judges for interpretation of the existence and identification of the source (Emitter) of utterances perceived in a natural speech situation. October, 1970.
3. SMASH: A Coding system for manuscript analysis of discrepancies among formal, grammatical and semantic strings covering the same oral stretch. (November, 1970).
4. A system for Minimal Control of Transliterations of Recorded Verbal Materials for use in Language Behavior Research.
5. The Uses of Transcripts in Educational Research.
6. The Uses of Transcripts in Psychological Research.
7. The Uses of Transcripts in Linguistic Research.
8. The Uses of Transcripts in Sociological Research.
9. The Uses of Transcripts in Courts of Law.
10. On the Non-identity of Speech and Writing.
11. Linguistics Illustrated: A new approach for explaining linguistics to the educated layman.

With Daniel G. Hays

1. Post-editing Procedures: A Working Manual. 22 April, 1969.

IN PREPARATION:
1. Cost analysis of Transcript Production. November, 1970.
2. Cost Analysis of Delayed Information Retrieval for Transcribing Effectiveness. December, 1970.

H. N. Hays, LSA, July 24, 1970

Copies of the first few lines of transcriptions

A, B, C, D of the same videotaped segment

// = Common point of origin. Parentheses enclose Emitter designates.
_____ = indistinguishable utterance.

Transcriber A

(T) All of us, yes. //Have all of you all completed the reading on page 109? All right let's discuss how a friendly letter can be like having a visit with someone. What are some of the that we a talk about in a friendly letter? What's the atmosphere of a friendly letter?
(?) You mean the parts.
(T) No not the parts, honey, the atmosphere. David.
(David) the attitude of the person who writes.

Transcriber B

(T) //Have all of you all completed the reading on page 109?
(Class) Yes
(T) Oh right let's discuss how a friendly letter can be like having a visit with someone. What are some of the things that we a talk about in a friendly letter? What's the atmosphere of a friendly letter?
(?) You mean the parts?
(T) No not the parts honey the atmosphere. Doyle?
(Doyle) You mean the attitude of the person like.

Transcriber C

(T) _____. //Have all of you completed the reading of page 109.
(Class) Murmurs.
(T) Allright, let's discuss how a formal letter can be like having a visit with someone. What are some of the things that we uh talk about in a formal letter? What's the atmosphere of a formal letter?
(?) _____
(T) No, not the parts, honey, the atmosphere, Dale.
(Dale) The attitude of the person writing it.

Transcriber D

(T) All right, have ya . . . Yes? //Have all of you all completed the reading on page 109?
(Class) Yes
(T) All right, let's discuss how a friendly letter, letter can be like having a visit with someone. What are some of the, uh, things that we, us, talk about in a friendly letter? What's the atmosphere of a friendly letter?
(?) The, uh, parts.
(T) No, not the parts, honey, the atmosphere. Dana?
(D) The attitude of the person you write.

TABLE III

Sample of Kinds of Discrepancies Tabulated for Transliterations Made by Four Different Transcribers of the Same Videotaped Segment

Discrepancies are circled
A blank space indicates no representation by transcriber
A line _____ indicates that transcriber could not distinguish the utterance

Category	(Cross Referenced)	Transcriptions			
		a	b	c	d
I. (P) Punctuation	1. (WI), (WO)	her O address at the top.	her (a-a-) address? at the top?	her O address at the top	her (,) uh (,) address at the top.
	2. (PI) (WI), (WO)	called? (W)hat is the greeting?	called? (O)r what is the greeting?	called? (O)r what is the greeting?	called? (W)hat is the greeting?
	3. (WI), (Sp)	oh. (Ah) June Butler.	oh? (Um-m-m-m) June O	(Oh?) (Uh,) .. June Butler.	oh? (Uh,) June Butler?
II. (WI) Word Interpretation	4. (none)	friendly letter	friendly letter	formal letter	friendly letter
	5. (Sp), (P)	(Allright.) Anything else	(Oh right.) Anything else.	(All right,) someone else.	(Alright,) anything else.
III. (WO) Word Omission	6. (P), (WI), (Sp)	letter to Teresa's grandfather, Shakinnea?	letter? to Teresa's Grandfather? Chickenya?	letter to Teresa's grandfather Shakineen?	letter to Teresa's grandfather? ( )
	7. (none)	You convey	You convey	You (can) convey	You convey
IV. (E) Emitter	8. (P), (Sp), (WI)	(David)	(Doyle?)	(Dale.)	(Dana?)
	9. (PI), (Sp), (PO), (SD), (WI)	(Shakinnea) Write a	(Chickenya does not answer)	( )	(?) (No answer)

TABLE III (continued)

Category	(Cross Referenced)	Transcriptions a	b	c	d
V. (PI) Phrase Interpretation	10. (WO), (P), (Sp)	(Class) She didn't capitalize East.	(Class) She didn't start it on the East.	(Class) Capital letter on east.	(Class) She didn't capitalize "East"
	11. (P), (WI)	Now we know why Billy is writing this up so high don't we?	Now we know why Bailey's writing is set up so high don't we?	Now we know why Billy's writing this up so high don't we?	Now, we know why Billy's writing this up so high, don't we?
	12. (P), (WO), (WI)	Some of the things that are experiences.	It tells of some experiences.	Things that tell experiences.	Uh, telling, telling things of, telling your experiences.
VI. (PO) Phrase Omission	13. (P), (WO), (PI)	_____	You write things you've done.	_____	You write the things you, the things you...
	14. (E), (SD)	(Bonnie) Thank you.	(To Bonnie)	(T)	
VII. (SP) Spelling	15. (none)	anything	anything	anythin	anything
	16. (WI), (PI), (P)	Are you going to cross that T?	Are you gonna cross that T?	You wanta cross that T? Thank you.	You gonna' cross that "t"?
VIII. (SD) Stage Directions	17. (none)		(Teresa writes on blackboard)	(Teresa corrects her error)	(Teresa corrects her mistake)
	18. (E), (WO)	Keel	_____ (Calls a name	Sheila?	Gail?

H.N. Hays, LSA, July 24, 1970

Multi-form Discrepancies

Sect. Num	Forms	Processor
83 I.	(David)	A
	(Doyle)	Bb;Br
	(Darla)	Fb,Ff2
	(Daryl)	Ff1,Ff2, Fh
	(Dale)	C,Bf,E
	(Dana)	Dd,Do,Dg
	(d)	D1
114 II.	Bryan	A;B
	Brian	D,F
	Frank	C
368 II.	Por	C,Fb,Fn,Bm1
	Por#	Ff,Fg,Bm2,E
	Port	D
	Port#	Af
	Portu	Bn
	%	R
399 II.	maybe	Fn
	%	Fb,Aa
	---	C
	later	R
490 II.	broad	Bn
	---	A,C
	%	D
	large	R
491-2, II.	in poverty and powerful	A Bm:Bn Bb,C,D

Sect. Num.	Forms	Processor
373 I.	Daryl	F
	Dale	Eg
	Gill	Eg,Ej
	Keel	A
	Gail	Bf,D,F Ee,Ee2 Bb

	Sheila	C
	Bill	Gn
448 II.	Harry	A
	Larry	Bb,Fb, Fn
	Mary	R
62 II.	so	A
	But	Bm,Bm2, Bn1,Bn2,F
	but	Bb,C,D
470 II.	has	F
	were	A;E
	%	A,B
	---	R
402 II.	man	Aa
	---	C
	land	R
243 II.	ruler	C,Bm,F, E
	unruley	Aa
	rulers	Af
	---	D,Bn,Bb
365 II.	used	D
	ruled	R
497 II.	ruled	B,F
	rural	D
	grew	A
	---	C

Sect. Num.	Forms	Processor
324 II.	Daryl	Bb;D,F
	Darryl	Bm,Bn,E
	Darrell	A,C
364 I.	(Shakina)	Fb,Fe,Ff1
	(Shekina)	Ff2
	(Shekeena)	Fh
	(Shakinnea)	E,A
	(S)	C
	(?)	Dd,Do
	%	B
336 II.	But	C
	but	F,B,E
	where	A
	%	D
32 II.	ready	C
	waiting	R
205 II.	portrait	Aa
	part	R
495 II.	influence	D,C
	emporer	Bn
	emperior	Fb,Fn
	emperor	R

(H.N. Hays, LSA, 24 July, 1970)

Figure 2
Example of Transcript Alignment for Analysis

| | | |

| | | |

H. N. Hays, LSA, July 24, 1970

GLOSSARY OF INFORMANTS AND PROCESSORS

Individual	Profession	Grade	Age	Color	Sex	Dialect: General	Dialect: Local	MS I	MS II	Experience	Rating	Function: Informant	Transcriber	Typist	Editor	Collator	Coder	Tabulator	Programmer	JK
E	T1	VI	Y	B	F	a	b u	bp	x	4	3	x								5
F	T2	VI	Y	W	F	a	b u	bp	x	4	3	x								5
C1	C	VI	C	B/W	M/F		b u	b	x		2	x								3
C2	C	VI	C	B/W	M/F		b u	b	x		2	x								3
a	S	H	Y	W	F		c t	c	x	x	4	3	x	x				x		3
b	S	H	Y	W	F		c r	d	x	x	3	3	x	x				x		3
c	S	4	Y	W	F	ga	cut	e	x	x	4	3	x	x				x		4
d	S	H	Y	W	F	a	c r	d	x	x	4	3	x	x				x		3
e	U	2	T	W	F		c r	d	x	x	3	2	x		x	x	x			3
f	N	3	Y	W	F	ga	cut		x	x	3	4			x					3
g	L	P	M	W	F	gy	cut	yp	x	x	5	5	x	x	x	x	x	x		5
h	LPG	D	M	W	M	gy	cut	xp			5	5	x	x	x			x	x	5
i	U	2	T	W	M	ga	cu	b			4	4						x	x	5
j	U	2	T	W	M	ga	cut	e			3	4	x		x					5
k	U	4	Y	B	F	gab	cy	bp		x	2	3			x		x			4
l	A	M	Y	W	F	ga	cut	cp			2	4			x					5
m	L	M	Y	W	F	ga	c u	mp	x	x	4	4		x		x	x	x		4
n	G	M	Y	W	F	ga	cut	sp	x	x	3	4		x		x	x	x		4
o	N	3	Y	W	F	ga	cut	c	x		1	2			x					1
p	G	M	Y	W	M	g	cut	kp	x		2	2			x					3
q	E	G	Y	W	M	ga	cur	c	x		1	2			x					2
r	U	F	Y	B	F	U	cut	u	x		1	3			x					1
s	U	4	Y	W	F	gc	gut	a	x	x	3	3			x	x	x			4
t	G	G	Y	W	M	ga	cuc	tp	x		1	2			x					1
u																				
v	L	M	M	W	F	gr	cut	n	x	x	5	5					x	x	x	5

Processors of Contracted Manuscripts (7 minutes of real time)

I

A	B	C	D	E	F	G
a	b	c	d	e	b	s
t1	r1	q1	o1	e1	e1	e1
t2	r2	q2	o2	g2	f2	e2
e1	f1	p1	ℓ1	j3	f3	
e2	f2	p2	ℓ2		n4	n3
			g1/2			

II

A	B	C	D	E	F	G
a1	b	c	d	s	b	s1
f2	m1			e1	e1	s2
	m2			e2	f2	e3
	n1				g3	e4
	n2				k4	e5
						e6

H. N. Hays, LSA, July 24, 1970

	I							II						
	A	B	C	D	E	F	G	A	B	C	D	E	F	G
1. Annotations	4	6	4	3		10	101	2	9	11	13		15	90
2. Universal	0	0	0	0	0	0								
3. Discrepant								7	14	11	17			
4. Omission	1	4	4	4										
5. Interpretation														
6. Tokens	9	9	9	9										
7. Emitter	51	52	58	58		26	80	39	31	34	38		27	54
8. U							29							12
9. D	6	2	8	5										42
10. O	5	2	5	4										
11. I	1	0	3	1										
12. Tokens	25	25	25	25										
13. Punctuation	84	83	87	81	81	67	81	67	71	73	74		74	183
14. U														30
15. D	12	13	14	13										153
16. O														
17. I														
18. Tokens	44	44	44	44										
19. Words	548	541	525	562	664	557	640	560	496	474	548		550	645
20. U							326							305
21. D							(47)							189
22. O	12	18	12	10										138
23. I	22	32	36	27										
24. Tokens	62	62	62	62										
25. Phrase							118						79	137
26. U							77							78
27. D	18						41							59
28. O	18													
29. Tokens	36	36	36	36										
30. Pages	2.2	2.2	2.5	2.5	(11)	4.5	87	4	5	7	5		4.5	118
31. Lines	76	72	81	84	(101)		(727)	71	61	63	76		63	(770)
32. Sweeps	5	5	5	7	(17)5	5	(32)2	2	5	1	1	(9)1	6	(16)4
33. Processors	3	3	3	4	(13)3	4	(20)4	2	3	1	1	(7)2	6	(10)2
34. T-E time	120	130	130	135	550	180		140"	695"	40	40	?	395	!
	A	B	C	D	E	F	G	A	B	C	D	E	F	G
				I							II			